

REC'D 29 APR 2004

WIPO

PCT

# THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

October 27, 2003

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: 60/463,747

FILING DATE: April 18, 2003

8c7IB04/1144

PRIORITY DOCUMENT  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH  
RULE 17.1(a) OR (b)



By Authority of the  
COMMISSIONER OF PATENTS AND TRADEMARKS

*P. R. Grant*

P. R. GRANT  
Certifying Officer

04/18/03



Please type a plus sign (+) inside this box →



04-21 1133 U.S. PTO 60/463747

PTO/SB/16 (02-01)

Approved for use through 10/31/2002. OMB 0651-0032  
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE  
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.**PROVISIONAL APPLICATION FOR PATENT COVER SHEET**

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53 (c).

Express Mail Label No. EL 827841893 US Date of Deposit: April 18, 2003

INVENTOR(S)					
Given Name (first and middle [if any])	Family Name or Surname		Residence (City and either State or Foreign Country)		
Jong Chul	Ye		Croton-on-Hudson, NY		
<input type="checkbox"/> Additional inventors are being named on the _____ separately numbered sheets attached hereto					
TITLE OF THE INVENTION (280 characters max)					
SYSTEM AND METHOD FOR RATE-DISTORTION OPTIMIZED DATA PARTITIONING FOR VIDEO CODING USING PARAMETRIC RATE-DISTORTION MODEL					
CORRESPONDENCE ADDRESS					
Direct all correspondence to:					
<input checked="" type="checkbox"/> Customer Number	24737		*24737*		
OR Type Customer Number here					
<input checked="" type="checkbox"/> Firm or Individual Name	PHILIPS ELECTRONICS NORTH AMERICA CORPORATION				
Address	580 WHITE PLAINS ROAD				
Address					
City	TARRYTOWN	State	NY	ZIP	10591
Country	USA	Telephone	(914) 333-9611	Fax	(914) 332-0615
ENCLOSED APPLICATION PARTS (check all that apply)					
<input checked="" type="checkbox"/> Specification Number of Pages	22	<input type="checkbox"/> CD(s), Number			
<input checked="" type="checkbox"/> Drawing(s) Number of Sheets	6	<input type="checkbox"/> Other (specify)			
<input type="checkbox"/> Application Data Sheet. See 37 CFR 1.76					
METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT (check one)					
<input type="checkbox"/> Applicant claims small entity status. See 37 CFR 1.27.					
<input type="checkbox"/> A check or money order is enclosed to cover the filing fees					
<input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: <u>14-1270</u> <u>160.00</u>					
<input type="checkbox"/> Payment by credit card. Form PTO-2038 is attached.					
The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.					
<input checked="" type="checkbox"/> No.					
<input type="checkbox"/> Yes, the name of the U.S. Government agency and the Government contract number are: _____					

Respectfully submitted,

SIGNATURE

Date April 18, 2003

TYPED or PRINTED NAME

EDWARD W. GOODMANREGISTRATION NO.: 28,613  
(if appropriate)

TELEPHONE

(914) 333-9611Docket Number: US030089**USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT**

This collection of information is required by 37 CFR 1.51. The information is used by the public to file (and by the PTO to process) a provisional application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 8 hours to complete, including gathering, preparing, and submitting the complete provisional application to the PTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, Washington, D.C., 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Box Provisional Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

1133 U.S. PTO  
60/463747  
04/18/03

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of

Atty. Docket

JONG CHUL YE

US030089

Serial No.

Group Art Unit:

Filed: CONCURRENTLY

Examiner:

Title: SYSTEM AND METHOD FOR RATE-DISTORTION OPTIMIZED DATA  
PARTITIONING FOR VIDEO CODING USING PARAMETRIC RATE-DISTORTION  
MODEL

Honorable Commissioner of Patents  
Washington, D.C. 20231

AUTHORIZATION PURSUANT TO 37 CFR 31.136(a)(3)  
AND TO CHARGE DEPOSIT ACCOUNT

Sir:

The Commissioner is hereby requested and authorized to treat any concurrent or future reply in this application requiring a petition for extension of time for its timely submission, as incorporating a petition for extension of time for the appropriate length of time.

Please charge any additional fees which may now or in the future be required in this application, including extension of time fees, but excluding the issue fee unless explicitly requested to do so, and credit any overpayment, to Deposit Account No. 14-1270.

Respectfully submitted,

By 

Edward W. Goodman, Reg. 28,613  
Attorney  
(914) 333-9611

# **SYSTEM AND METHOD FOR RATE-DISTORTION OPTIMIZED DATA PARTITIONING FOR VIDEO CODING USING PARAMETRIC RATE- DISTORTION MODEL**

5

## **BACKGROUND OF THE INVENTION**

### **1. Field Of The Invention**

10 The present invention is related to scalable video coding systems, in particular, the invention relates to a general rate-distortion optimized data partitioning (gRDDP) of discrete cosine transform (DCT) coefficients for video transmission over packet lossy network using a parametric rate-distortion (RD) model.

### **2. Description Of The Related Art**

15 Video is a sequence of pictures; each picture is formed by an array of pixels. The size of uncompressed video is huge. To reduce its size, video compression may be used to reduce the size and improve the data transmission rate. Various video coding methods (e.g., MPEG 1, MPEG 2, and MPEG 4) have been established to provide an international standard for the coded representation of moving pictures and associated audio on digital storage media.

20 Such video coding methods format and compress the raw video data for reduced rate transmission. For example, the format of the MPEG 2 standard consists of 4 layers: Group of Pictures, Pictures, Slice, Macroblock. A video sequence begins with a sequence header that includes one or more groups of pictures (GOP), and ends with an end-of- sequence code. The Group of Pictures (GOP) includes a header and a series of one of more pictures intended to allow random access into the video sequence.

25 The pictures are the primary coding unit of a video sequence. A picture consists of three rectangular matrices representing luminance (Y) and two chrominance (Cb and Cr) values. The Y matrix has an even number of rows and columns. The Cb and Cr matrices are one-half the size of the Y matrix in each direction (horizontal and vertical). The slices are one or more "contiguous" macroblocks. The order of the macroblocks within a slice is from left-to-right and top-to-bottom.

703553

The macroblocks are the basic coding unit in the MPEG algorithm. The macroblock is a 16x16 pixel segment in a frame. Since each chrominance component has one-half the vertical and horizontal resolution of the luminance component, a macroblock consists of four Y, one Cr, and one Cb block. The Block is the smallest coding unit in the MPEG algorithm. It consists of 8x8 pixels and can be one of three types: luminance(Y), red chrominance (Cr), or blue chrominance(Cb). The block is the basic unit in intra frame coding.

The MPEG 2 standard defines three types of pictures: Intra Pictures (I-Pictures) Predicted Pictures (P-Pictures); and Bidirectional Pictures (B-Pictures). Intra pictures, or I-Picture, are coded using only information present in the picture itself, and provides potential random access points into the compressed video data. Predicted pictures, or P-pictures, are coded with respect to the nearest previous I- or P-pictures. Like I-pictures, P-pictures also can serve as a prediction reference for B-pictures and future P-pictures. Moreover, P-pictures use motion compensation to provide more compression than is possible with I-pictures. Bidirectional pictures, or B-pictures, are pictures that use both a past and future picture as a reference. B-pictures provide the most compression since it uses the past and future picture as a reference. These three types of pictures are combined to form a group of picture.

The MPEG transform coding algorithm includes the following coding steps:  
Discrete cosine transform (DCT), Quantization, and Run-length encoding .

An important technique in video coding is scalability. In this regard, a scalable video codec is defined as a codec that is capable of producing a bitstream that can be divided into embedded subsets. These subsets can be independently decoded to provide video sequences of increasing quality. Thus, a single compression operation can produce bitstreams with different rates and reconstructed quality. A small subset of the original bitstream can be initially transmitted to provide a base layer quality with extra layers subsequently transmitted as enhancement layers. Scalability is supported by most of the video compression standards such as MPEG-2, MPEG-4 and H.263.

An important application of scalability is in error resilient video transmission. Scalability can be used to apply stronger error protection to the base layer than to the enhancement layers (i.e., unequal error protection). Thus, the base layer will be

successfully decoded with high probability even during adverse transmission channel conditions.

Data Partitioning (DP) is used to facilitate scalability. For example in MPEG 2, the slice layer indicates the maximum number of block transform coefficients contained in the particular bitstream (known as the priority break point). Data partitioning is a frequency domain method that breaks the block of 64 quantized transform coefficients into two bitstreams. The first, higher priority bitstream (e.g., base layer) contains the more critical lower frequency coefficients and side information (such as DC values, motion vectors). The second, lower priority bitstream (e.g., enhancement layers) carries higher frequency AC data.

Figure 1 shows a block diagram illustrating data partitioning that may be implemented outside the encoder. At the transmitter, the demultiplexer receives from the variable length decoder (VLD) the number of bits used for each variable length code and separates the bitstream based on the priority break point (PBP) value. Note that the PBP's can be changed at each slice based on the rate partitioning logic used. In particular, in conventional DP video coders (e.g., MPEG), single layer bit stream is partitioned into two or more bit streams in the DCT domain. During transmission, one or more bit streams are sent to achieve bit rate scalability. Unequal error protection can be applied to base and enhancement layer data to improve robustness to channel degradation.

Figure 2 shows a block diagram illustrating merging that may be implemented outside the decoder. As shown, two VLD's are used to process the base layer and enhancement layer streams and then output nonlayered bitstream. The PBP defines how an encoded bitstream is partitioned. Before decoding, depending on resource allocation and/or receiver capacity, the received bitstreams or a subset of them are merged into one single bitstream and decoded.

The conventional DP structure has advantages in a home network environment. More specifically, at its full quality, the rate-distortion performance of the DP is as good as its single layer counterpart while rate scalability is also allowed. The rate-distortion (R-D) performance is concerned with finding an optimal combination of rate and distortion. This optimal combination, which could also be seen as the optimal combination of cost and quality, is not unique. R-D schemes attempt to represent a piece of information with

the fewest bits possible and at the same time in a way that will lead to the best reproduction quality.

It is also noted that in the conventional DP structure, the additional decoding complexity overhead is very minimal at its full quality while the DP provides wider range of decoder complexity scalability. This is because variable length decoding (VLD) of DCT run-length pairs – which is the most computational extensive part – now becomes scalable.

In the conventional DP structure, the DCT priority break point (PBP) value needs to be transmitted explicitly as side information. To minimize the overhead, the PBP value is usually fixed for all the DCT blocks within each slice or video packet.

While the conventional DP method is simple and has some advantages, it is not capable of adapting base layer optimization because only one PBP value is used for all blocks within each slice or video packets. In addition, a prediction drift occurs at low bit rates as a result of the single-loop prediction structure used for data partitioning. Thus, it is difficult during data partitioning how to choose the DCT break point for each block such that the base station quality at a given base partition rate is optimal. In order to achieve a minimum distortion at the base layer, the partitioning point must be allowed to vary at the DCT block level. However, such a fine control of the breakpoint introduces significant rate overhead due to the explicit transmission of breakpoint values.

Accordingly, there exists a need for video coding techniques that overcome the limitations of the conventional data partitioning scheme and provide improved base layer optimization.

## SUMMARY OF THE INVENTION

The present invention addresses the foregoing need and provides additional advantages, by providing an improved data partitioning technique by employing a parametric RD model. In one embodiment of the present invention, this can be achieved with minimal overhead ( $\approx 20$  bits for each slice or video packet or even for each frame) by employing context-based backward adaptation.

703553

One aspect of the present invention is directed to a system and method that provide a rate-distortion optimized data partitioning (gRD-DP) of DCT coefficients for video transmission.

In another aspect of the present invention, the RD-DP adapts the partition point  
 5 block-by-block, hence greatly improves the coding efficiency of the base layer bit stream. This also allows a decoder to find the partition location in backward-fashion from the decoded data without explicit transmission, hence saving the bandwidth significantly.

In yet another aspect of the present invention, a Lagrangian parameter  $\lambda$  is calculated. The value of  $\lambda$  is determined to meet the rate budget  $R_b$  (for the base layer  
 10 transmission channel) using a standard one-dimensional bisection algorithm.

One embodiment of the present invention is directed to a data partitioning method for a scalable video encoder. The method includes the steps of receiving video data; determining DCT coefficients for a plurality of macroblocks of a video frame; quantizing the DCT coefficients and converting the quantized DCT coefficients into (run, length)  
 15 pairs; determining the slope of the parametric rate-distortion curve for each the plurality of macroblocks in the video frame, wherein if the slope is less than  $\lambda$  or if the k-th slope is a first slope that is not less than  $\lambda$ , write the k-th (run, length) pair into the base layer, otherwise if the k-th slope is greater than  $\lambda$ , write the k-th (run, length) pair into the at least one enhancement layer, where  $\lambda$  is determined in accordance with a Lagrangian  
 20 calculation.

Another embodiment of the present invention is directed to a method for determining a boundary between a base layer and at least one enhancement layer in a scalable video decoder. The method includes the steps of receiving the base layer and the at least one enhancement layer, the base layer and enhancement layer including data  
 25 representing (run, length) pairs for a plurality of macroblocks in a video frame. For each the plurality of macroblocks in the video frame, determining the slope of the parametric rate-distortion curve. If the slope is less than  $\lambda$  or if the k-th slope is a first slope that is not less than  $\lambda$ , read the k-th (run, length) pair from the base layer, otherwise if the k-th slope is greater than  $\lambda$ , read the k-th (run, length) pair from the at least one enhancement  
 30 layer, where  $\lambda$  is determined in accordance with a Lagrangian calculation.

Yet another embodiment of the present invention is directed to a scalable decoder



capable of merging data from a base layer and at least one enhancement layer. The decoder includes a memory which stores computer-executable process steps, and a processor which executes the process steps stored in the memory so as (i) receiving the base layer and the at least one enhancement layer, the base layer and enhancement layer including data representing (run, length) pairs for a plurality of macroblocks in a video frame, (2) for each the plurality of macroblocks in the video frame, determining a parametric rate-distortion model, (3) computing the slope (tangent) of the parametric rate-distortion model at using k (run,length) pairs, for an i-th block, and (3) if the slope of the parametric model updated using k (run,length) pairs is less than  $\lambda$  or if the it is a first slope that is not less than  $\lambda$ , read the k-th (run, length) pair from the base layer, otherwise if the the slope is greater than  $\lambda$ , read the k-th (run, length) pair from the at least one enhancement layer, where  $\lambda$  is determined in accordance with a Lagrangian calculation.

Yet another embodiment of the present invention is directed to a scalable transcoder. A single layer coded video bitstream (MPEG-1, MPEG-2, MPEG-4, H.264, etc) is partially decoded and the bitstream splitting point is determined for each DCT block based on the forementioned boundary determining method embodiment. Afterwards the VLC codes are split into two or more partitions based on the splitting points. The partial decoding involves variable length decoding, inverse scanning and inverse quantization only. No inverse DCT or motion compensation is needed.

The invention has particular utility in connection with variable-bandwidth networks and computer systems that are able to accommodate different bit rates, and hence different quality images.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1 and 2 are general block diagrams of a system for data partitioning and merging.

Figure 3 depicts a video coding system in accordance with one aspect of the present invention.

Figure 4 depicts a typical convex Rate-Distortion curve.

Figure 5 depicts a non-convex Rate-Distortion curve.

Figure 6 depicts a computer system on which the present invention may be implemented.

Figure 7 depicts the architecture of a personal computer in the computer system shown in Figure 6.

Figure 8 depicts a block diagram of a transcoder in accordance with one embodiment of the present invention.

# DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

10

Figure 3 illustrates a scalable video system 100 with layered coding and transport prioritization. A layered source encoder 110 encodes input video data. The output of the layered source encoder 110 includes a base layer 121 and one or more enhancement layers 122-124. A plurality of channels 120 carry the output encoded data. A layered source decoder 130 decodes the encoded data.

There are different ways of implementing layered coding. For example, in temporal domain layered coding, the base layer contains a bit stream with a lower frame rate and the enhancement layers contain incremental information to obtain an output with higher frame rates. In spatial domain layered coding, the base layer codes the sub-sampled version of the original video sequence and the enhancement layers contain additional information for obtaining higher spatial resolution at the decoder.

Generally, a different layer uses a different data stream and has distinctly different tolerances to channel errors. To combat channel errors, layered coding is usually combined with transport prioritization so that the base layer is delivered with a higher degree of error protection. If the base layer 121 is lost, the data contained in the enhancement layers 122-124 may be useless.

In one embodiment of the present invention, the video quality of the base layer 121 is flexibly controlled at the DCT block level. The desired base layer can be controlled by adapting the break points at the DCT block level by employing parametric RD model to approximate the convex hull of the RD planes for each DCT blocks, thereby finding the optimal partitioning points synchronously at the encoder and decoder (explained later with

703553

reference to Figures 5 and 6).

It is noted that the purpose of DCT is to reduce the spatial correlation between adjacent error pixels, and to compact the energy of the error pixels into a few coefficients. Because many high frequency coefficients are zero after quantization, variable length coding (VLC) is accomplished by a *runlength coding* method, which orders the coefficients into a one-dimensional array using a so-called zig-zag scan so that the low-frequency coefficients are put in front of the high-frequency coefficients. This way, the quantized coefficients are specified in terms of the non-zero values and the number of the preceding zeros. Different symbols, each corresponding to a pair of zero runlength, and non-zero value, are coded using variable length codewords.

The scalable video system 100 preferably uses entropy coding. In entropy coding, quantized DCT coefficients are rearranged into a one-dimensional array by scanning them in a zig-zag order. This rearrangement puts the DC coefficient at the first location of the array and the remaining AC coefficients are arranged from the low to high frequency, in both the horizontal and vertical directions. The assumption is that the quantized DCT coefficients at higher frequencies would likely be zero, thereby separating the non-zero and zero parts. The rearranged array is coded into a sequence of the run-level pair. The *run* is defined as the distance between two non-zero coefficients in the array. The *level* is the non-zero value immediately following a sequence of zeros. This coding method produces a compact representation of the 8 x 8 DCT coefficients, since a large number of the coefficients have been already quantized to zero value.

The run-level pairs and the information about the macroblock, such as the motion vectors, and prediction types, are further compressed using entropy coding. Both variable-length and fixed-length codes are used for this purpose.

The design of the video system 100 is motivated by the operational rate-distortion (RD) theory. RD theory is useful in coding and compression scenarios, where the available bandwidth is known a priori and where the purpose is to achieve the best reproduction quality that can be achieved within this bandwidth (i.e., adaptive algorithms).

Discussed below is an illustration formulated to solve for the optimized partitions (i.e., base and enhancement layer partitions). In the following discussion it is assumed

that there are “n” DCT blocks for each video frame and the bit rate budget  $R_b$  is known for the base layer partition. The rate budget is determined based on the minimal video quality requirement and channel throughput fluctuation. Then, the following optimization problem can be formulated to solve for the optimal partitions:

5

$$\min_{P_1, \dots, P_n} \sum_{i=1}^n Di(P_i) \quad \text{subject to} \quad \sum_{i=1}^n Ri(P_i) \leq R_b \quad (1)$$

where  $P_i \in \{0, 1, \dots, K(i)\}$ ,  $i = 1, \dots, n$  is the break point value for the  $i$ -th block and  $K(i)$  denotes the maximum (run, length) pairs in the  $i$ -th block,  $R_i(P_i)$  and  $D_i(P_i)$  denote the corresponding bit rate and the distortion from the  $i$ -th block, respectively.

10

The optimization problem can be solved using an iterative bisection algorithm based on a Lagrangian optimization. The optimal partitioning point  $P_i$  satisfies the following condition for all  $i = 1, \dots, n$ :

$$\frac{\partial Di(P_i)}{\partial Ri(P_i)} + \lambda = 0, \quad i = 1, \dots, n \quad (2)$$

15

where the Lagrangian  $\lambda > 0$  is determined by the standard bisection search so that the rate constraint in (1) is satisfied.

If the  $k$ -th DCT (run, length) pair for the  $i$ -th block is  $L_i^k$  bits and has a coefficient value of  $X_i^k$ ; then, the slope for the rate-distortion (R-D) curve of the  $i$ -th block at the  $k$ -th DCT (run, length) pair has the following set of discrete values:

20

$$\frac{\partial Di(P_i)}{\partial Ri(P_i)} = \frac{Di(P_{i+1}) - Di(P_i)}{Ri(P_{i+1}) - Ri(P_i)} \in \left\{ \frac{|X_i^k|^2}{L_i^k} \right\}_{k=1}^{K(i)} \quad (3)$$

25

Referring now to Fig. 4, a convex R-D curve is shown to illustrate how to determine the partition point and how the layered source decoder 130 can infer the partition point in a backward-adaptive fashion. It is noted that the layered source decoder 130 operates in the same way even if R-D curve is not convex.

From Fig. 4, if the rate-distortion curve is convex it can be seen that in general  $\lambda$  is a decreasing function with respect to R and therefore, in general, the following relationship holds:

$$\frac{|X_i^1|^2}{L_i^1} \geq \frac{|X_i^2|^2}{L_i^2} \geq \dots \geq \frac{|X_i^{K(i)}|^2}{L_i^{K(i)}} \quad (4)$$

In accordance with Eq. (4) a partitioning algorithm for the DCT coefficients at the layered source encoder 110 side is given below if the rate-distortion curve is convex. It is noted that to get to this point, the video data for a frame is converted using the *discrete cosine transform* (DCT), the DCT coefficients are quantized, and then converted into binary codewords (run, length) using *variable length coding* (VLC).

```

for i=1,...,n    { for each macroblock in frame
                  for k=1,...,K(i) { for each (run, length) pair
15                                     Compute the corresponding  $X_i^k, L_i^k$ .
                                     Put the k-th (run, length) VLC into base layer.
                                     if  $|X_i^k|^2 / L_i^k < \lambda$  break;
                                     }
                  put the remaining (run, length) pairs of i-th block into ENH layer.
20      }

```

The Lagrangian parameter  $\lambda$  may be separately encoded and transmitted as side information (i.e., overhead information). The layered source decoder 130 can find the boundary of the base layer 121 and enhancement layer 122, as well as, find the synchronization using the following algorithm:

```

for i=1,...,n    { for each macroblock in frame
                  for k=1,...,K(i) { for each (run, length) pair
30                                     Read VLC (run, length) pair from base layer.
                                     Compute the corresponding  $X_i^k, L_i^k$ .

```

if  $|X_i^k|^2 / L_i^k < \lambda$  break;

}

Read the remaining (run, length) pairs of I-th block from ENH layer.

}

5

As discussed above, the only side information to be transmitted is the Lagrangian parameter  $\lambda$ . The value of  $\lambda$  is determined to meet the rate budget Rb of Eq.(1) using the standard the one-dimensional bisection algorithm. However, the optimal value of  $\lambda$  can be a real number and should be quantized for transmission over the channel 120.

10

In practical implementation of variable length coding for the (run, length) pair, however, the R-D curve of Fig. 4 may be non-convex, as shown in Fig. 5, as the VLC is only an approximation of the true entropy of the source. In that case, the test variable  $|X_i^k|^2 / L_i^k$  is no longer monotonic with respect to k. In this case, the partitioning rule given by Eq.(4) is not valid and the near-optimality of RDDP can be broken, as shown in FIG. 5. Note that the optimal breakpoint value may be  $k_2$  while the RDDP algorithm provides  $k_1$ , which makes the base layer under-partitioned.

15

Accordingly, in a preferred embodiment, the convex hull is approximated using a parametric model which is continuously being updated at the encoder and decoder simultaneously using previously decoded (run, length) pairs.

20

More specifically, in a preferred embodiment, the following partitioning rule:

$$\left| \frac{\partial D_i(R_i(k); \theta_i(k))}{\partial R_i(k)} \right| \begin{cases} > \lambda & , \quad k \leq B_i \\ \leq \lambda & , \quad k > B_i \end{cases} \quad (5)$$

25

where  $D_i(R; \theta)$  denotes the i-th block base layer distortion model with respect to the rate R with a parameter vector  $\theta_i$ ,  $R_i(k)$  denotes the rate if k-(run, level) pairs are included, and  $\theta_i(k)$  is an estimated parameter for the i-th block using k-(run, level) pairs.

In Eq.(5), any rate distortion model can be used as long as it is convex and monotonically decreasing function. For example, an exponential distortion model may be used:

30

$$D(R; \theta) = \sigma^2 \exp(-\alpha R) \quad (6)$$

where  $\theta = (\sigma, \alpha)$  is the unknown parameter vector to be estimated.

5

For the distortion model Eq.(6), the partitioning rule becomes:

$$\sigma^2(k) \alpha(k) \exp(-\alpha(k) R_i(k)) \begin{cases} > \lambda & , k \leq B_i \\ \leq \lambda & , k > B_i \end{cases}$$

10

where  $\sigma(k), \alpha(k)$  are estimated parameter using the k-(run,level) VLC pairs.

Accordingly, the layered source decoder 130 can find the boundary of the base layer 121 and enhancement layer 122, as well as, find the synchronization using the following algorithm to split the bit-stream nearly optimally without sending explicit information of the  
15 breakpoint values:

#### Encoding:

20

Encode  $\lambda$  into base partition.

for I=1,...,N { // for each DCT blocks

for k=1,...,K(I) { //for each (run,level) pair

Compute  $\hat{C}_i(k)$  and  $L_i(k)$ .

25

Estimate  $\theta_i(k)$  using  $\{\hat{C}_i(m)\}_{m=1}^k$  and  $\{L_i(m)\}_{m=1}^k$  and update the parametric distortion function  $Di(R_i(k), \theta_i(k))$

Put the k-th (run,level) VLC into base partition.

If  $\left| \frac{\partial D_i(R_i(k); \theta_i(k))}{\partial R_i(k)} \right| < \lambda$  break.

end

30

Put the remaining (run, level) pairs into enhancement partition.  
end

**Decoding:**

Decode  $\lambda$  from base partition.  
 for I=1,...,N { // for each DCT blocks  
 5     for k=1,...,K(I) { //for each (run,level) pair  
        Read the k-th (run,level) VLC from base partition.  
        Compute  $\hat{C}_i(k)$  and  $L_i(k)$ .  
        Estimate  $\theta_i(k)$  using  $\{\hat{C}_i(m)\}_{m=1}^k$  and  $\{L_i(m)\}_{m=1}^k$  and update the  
        parametric distortion function  $Di(R_i(k), \theta_i(k))$   
 10         If  $\left| \frac{\partial D_i(R_i(k); \theta_i(k))}{\partial R_i(k)} \right| < \lambda$  break.  
        end  
        Read the remaining (run, level) pairs from enhancement partition.  
 15     end

As explained above, the only side information to be transmitted is the Lagrangian parameter  $\lambda$ . The value of  $\lambda$  is determined to meet the rate budget Rb of Eq.(1) using the standard the one-dimensional bisection algorithm. Then, it is quantized and transmitted once  
 20 for each frame header, hence the rate overhead is negligible.

Therefore, by transmitting the  $\lambda$  value and the corresponding low frequency and some high frequency DCT coefficients (as the base layer 121) over a more reliable transmission channel, greater dynamic allocation of the DCT information is achievable. This allows for more control of the minimal quality of the video in case data from one or  
 25 more of the enhancement layers 122-124 is lost.

Furthermore, the parametric model approximates the convex hull of the rate distortion curve, hence preventing under-partitioning from occurring even in non-convex rate-distortion function cases.

The embodiments of the present invention discussed above are applicable to any  
 30 scalable video coding system, e.g., MPEG 2, MPEG 4, H.263, etc.

Figure 6 shows a representative embodiment of a computer system 9 on which the present invention may be implemented. As shown in Figure 3, personal computer ("PC") 10 includes network connection 11 for interfacing to a network, such as a variable-bandwidth network or the Internet, and fax/modem connection 12 for interfacing with



other remote sources such as a video camera (not shown). PC 10 also includes display screen 14 for displaying information (including video data) to a user, keyboard 15 for inputting text and user commands, mouse 13 for positioning a cursor on display screen 14 and for inputting user commands, disk drive 16 for reading from and writing to floppy disks installed therein, and CD-ROM drive 17 for accessing information stored on CD-ROM. PC 10 may also have one or more peripheral devices attached thereto, such as a scanner (not shown) for inputting document text images, graphics images, or the like, and printer 19 for outputting images, text, or the like.

Figure 7 shows the internal structure of PC 10. As shown in Figure 7, PC 10 includes memory 20, which comprises a computer-readable medium such as a computer hard disk. Memory 20 stores data 23, applications 25, print driver 24, and operating system 26. In preferred embodiments of the invention, operating system 26 is a windowing operating system, such as Microsoft Windows2000; although the invention may be used with other operating systems as well. Among the applications stored in memory 20 are scalable video coder 21 and scalable video decoder 22. Scalable video coder 21 performs scalable video data encoding in the manner set forth in detail below, and scalable video decoder 22 decodes video data that has been coded in the manner prescribed by scalable video coder 21.

Also included in PC 10 are display interface 29, keyboard interface 30, mouse interface 31, disk drive interface 32, CD-ROM drive interface 34, computer bus 36, RAM 37, processor 38, and printer interface 40. Processor 38 preferably comprises a microprocessor or the like for executing applications, such those noted above, out of RAM 37. Such applications, including scalable video coder 21 and scalable video decoder 22, may be stored in memory 20 (as noted above) or, alternatively, on a floppy disk in disk drive 16 or a CD-ROM in CD-ROM drive 17. Processor 38 accesses applications (or other data) stored on a floppy disk via disk drive interface 32 and accesses applications (or other data) stored on a CD-ROM via CD-ROM drive interface 34.

Application execution and other tasks of PC 4 may be initiated using keyboard 15 or mouse 13, commands from which are transmitted to processor 38 via keyboard interface 30 and mouse interface 31, respectively. Output results from applications running on PC 10 may be processed by display interface 29 and then displayed to a user

on display 14 or, alternatively, output via network connection 11. For example, input video data which has been coded by scalable video coder 21 is typically output via network connection 11. On the other hand, coded video data received from, e.g., a variable bandwidth-network is decoded by scalable video decoder 22 and then displayed  
5 on display 14. To this end, display interface 29 preferably comprises a display processor for forming video images based on decoded video data provided by processor 38 over computer bus 36, and for outputting those images to display 14. Output results from other applications, such as word processing programs, running on PC 10 may be provided to printer 19 via printer interface 40. Processor 38 executes print driver 24 so as to perform  
10 appropriate formatting of such print jobs prior to their transmission to printer 19.

Another embodiment of the present invention is directed to a scalable transcoder. As shown in Fig. 8, a single layer coded video bitstream 200 (MPEG-1, MPEG-2, MPEG-4, H.264, etc) is partially decoded by a variable length decoder 210. The DCT coefficient  
15 220 are sent to an inverse scan/quantization unit 230 and then to a partitioning line finder 240. The bitstream splitting point is determined for each DCT block based on the boundary determining method embodiment discussed above. Afterwards VLC codes 250 are split into two or more partitions based on the splitting points. The results are provided to a variable length code buffer 260. In accordance with the embodiment, the partial  
20 decoding involves variable length decoding, inverse scanning and inverse quantization only. No inverse DCT or motion compensation is needed

Although the embodiments of the invention described herein are preferably implemented as computer code, all or some of the embodiments discussed above can be implemented using discrete hardware elements and/or logic circuits. Also, while the encoding and decoding techniques of the present invention have been described in a PC  
25 environment, these techniques can be used in any type of video devices including, but not limited to, digital televisions/set top boxes, video conferencing equipment, and the like.

In this regard, the present invention has been described with respect to particular illustrative embodiments. For example, principles of the present invention as described in the embodiments above may also be applied to partition enhancement layers. It is to be  
30 understood that the invention is not limited to the above-described embodiments and modifications thereto, and that various changes and modifications may be made by those

703553

of ordinary skill in the art without departing from the spirit and scope of the appended claims.

WHAT IS CLAIMED IS:

1. A method for partitioning data for a scalable video encoder, the method comprising the steps of:
  - 5 receiving video data;
  - determining DCT coefficients for a plurality of macroblocks of a video frame;
  - quantizing the DCT coefficients;
  - converting the quantized DCT coefficients into (run, length) pairs; and
  - 10 for each the plurality of macroblocks in the video frame, determining a ratio  $\left| \frac{\partial D_i(R_i(k); \theta_i(k))}{\partial R_i(k)} \right|$ , where a  $D_i(R; \theta)$  represents a distortion model for an i-th block,  $R_i(k)$  represents a rate for a k-(run, level) pair, and  $\theta_i(k)$  represents an estimated parameter for the i-th block using a k-(run, level) pair, and
    - if  $\left| \frac{\partial D_i(R_i(k); \theta_i(k))}{\partial R_i(k)} \right|$  is less than  $\lambda$  or if  $\left| \frac{\partial D_i(R_i(k); \theta_i(k))}{\partial R_i(k)} \right|$  is a first
    - 15 ration that is not less than  $\lambda$ , putting the k-th (run, length) pair into a base layer,
    - otherwise if  $\left| \frac{\partial D_i(R_i(k); \theta_i(k))}{\partial R_i(k)} \right|$  is greater than  $\lambda$ , putting the k-th (run, length) pair into an enhancement layer, where  $\lambda$  is determined in accordance with a Lagrangian calculation.
- 20 2. The method according to Claim 1, further comprising the step of transmitting the base and enhancement layers over different transmission channels.
3. The method according to Claim 1, wherein scalable video encoder
- 25 is an MPEG 4 encoder.

4. The method according to Claim 1, wherein scalable video encoder is an H.263 encoder.

5. The method according to Claim 1, wherein scalable video encoder is an MPEG 2 encoder.

6. The method according to Claim 1, wherein scalable video encoder is a video encoder which has DCT transform and entropy coding.

7. The method according to Claim 1, wherein scalable video encoder is realized by transcoding single layer MPEG2, MPEG4, and H.26L.

8. The method according to Claim 1, further comprising the step of quantizing  $\lambda$  and transmitting the quantized value as side information to a decoder.

9. The method according to Claim 6, wherein the side information is sent only once in a frame header for the video frame.

10. The method according to Claim 6, wherein the side information can be sent to a slice header or a video packet header to improve robustness.

11. The method according to Claim 1, wherein  $\lambda$  is determined to meet a rate budge for a transmission channel for the base layer using a bisection algorithm.

12. The method according to Claim 1, wherein  $\lambda$  is determined to meet a rate budge for a transmission channel for the base layer using an adaptive algorithm.

13. A method for determining a boundary between a base layer and at least one enhancement layer in a scalable video decoder, the comprising the steps of:

receiving the base layer and the at least one enhancement layer, the base layer and enhancement layer including data representing (run, length) pairs for a plurality of macroblocks in a video frame;

for each the plurality of macroblocks in the video frame, determining a ratio

$$\left| \frac{\partial D_i(R_i(k); \theta_i(k))}{\partial R_i(k)} \right|, \text{ where } D_i(R; \theta) \text{ represents a distortion model for an } i\text{-th}$$

block,  $R_i(k)$  represents a rate for a  $k$ -(run, level) pair, and  $\theta_i(k)$  represents an estimated parameter for the  $i$ -th block using a  $k$ -(run, level) pair, and

$$\text{if } \left| \frac{\partial D_i(R_i(k); \theta_i(k))}{\partial R_i(k)} \right| \text{ is less than } \lambda \text{ or if } \left| \frac{\partial D_i(R_i(k); \theta_i(k))}{\partial R_i(k)} \right| \text{ is the first}$$

ration that is not less than  $\lambda$ , read the  $k$ -th (run, length) pair from the base layer,

otherwise if the ratio  $\left| \frac{\partial D_i(R_i(k); \theta_i(k))}{\partial R_i(k)} \right|$  is greater than  $\lambda$ , read the  $k$ -th (run,

length) pair from the at least one enhancement layer, where  $\lambda$  is determined by decoding side information.

14. The method according to Claim 13, further comprising the step of receiving the base layer and enhancement layer over different transmission channels.

15. The method according to Claim 13, wherein scalable video decoder in an MPEG 4 decoder.

16. The method according to Claim 13, wherein scalable video decoder in an H.263 decoder.

17. The method according to Claim 13, wherein scalable video decoder in an MPEG 2 decoder.

18. The method according to Claim 13, wherein scalable video decoder in a video decoder that uses DCT and entropy coding.

19. The method according to Claim 13, wherein scalable video decoder is realized by a merger in front of a single layer video decoder selected from the group consisting of an MPEG2, MPEG4, and H.26L decoder.

20. The method according to Claim 13, further comprising the step of receiving  $\lambda$  as side information associated with the video frame.

21. The method according to Claim 20, wherein the side information is sent only once in a frame header for the video frame.

22. The method according to Claim 20, wherein the side information is copied for each slice header or video packet header to increase robustness.

23. The method according to Claim 13, wherein  $\lambda$  is determined to meet a rate budge for a transmission channel for the base layer.

24. A scalable decoder capable of merging data from a base layer and at least one enhancement layer, comprising:

- a memory which stores computer-executable process steps; and
- a processor which executes the process steps stored in the memory so as (i) receiving the base layer and the at least one enhancement layer, the base layer and enhancement layer including data representing (run, length) pairs for a plurality of macroblocks in a video frame, and (2) for each the plurality of macroblocks in the video frame, determining a ratio  $\left| \frac{\partial D_i(R_i(k); \theta_i(k))}{\partial R_i(k)} \right|$ , where a  $D_i(R; \theta)$  represents a distortion model for an i-th block,  $R_i(k)$  represents a rate for a k-(run, level) pair, and  $\theta_i(k)$  represents an estimated parameter for the i-th block using a k-(run, level) pair, and (3) if  $\left| \frac{\partial D_i(R_i(k); \theta_i(k))}{\partial R_i(k)} \right|$  is less than  $\lambda$  or if  $\left| \frac{\partial D_i(R_i(k); \theta_i(k))}{\partial R_i(k)} \right|$  is a first ratio that is not less than  $\lambda$ , read the k-th (run, length) pair from the base layer, otherwise if  $\left| \frac{\partial D_i(R_i(k); \theta_i(k))}{\partial R_i(k)} \right|$  is greater than  $\lambda$ , read the k-th (run, length) pair from the at least one enhancement layer, where  $\lambda$  is determined in accordance with a Lagrangian calculation.

25. The decoder according to Claim 24, wherein  $\lambda$  is received by the decoder as side information associated with the video frame and the side information is sent only once in a frame header for the video frame.

26. The decoder according to Claim 24, wherein  $\lambda$  is determined to meet a rate budge for a transmission channel for the base layer.



# ABSTRACT

A system and method are disclosed that provide a simple and efficient layered video coding technique using a parametric rate-distortion (RD) model. The video coding system may include an rate-distortion optimized data partitioning encoder and decoder. The generalized RD-DP encoder adapts the partition point block-by-block which greatly improves the coding efficiency of the base layer bit stream without explicit transmission thereby saving the bandwidth significantly. Furthermore, even for the non-parametric rate-distortion curves, the parameteric rate-distortion model prevents the underpartitioning of the base-layer from happening, and the parametric model is simultaneously being updated at the encoder and decoder for synchronization.

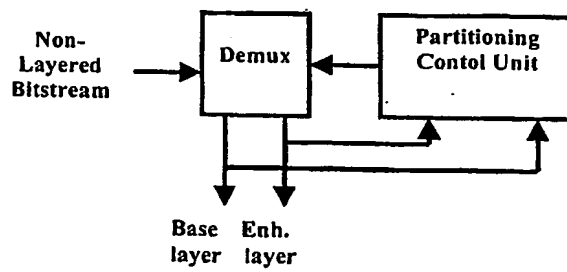


Figure 1

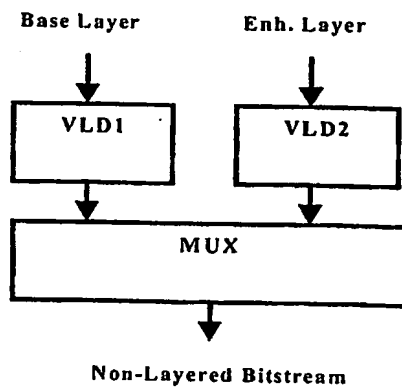


Figure 2

100

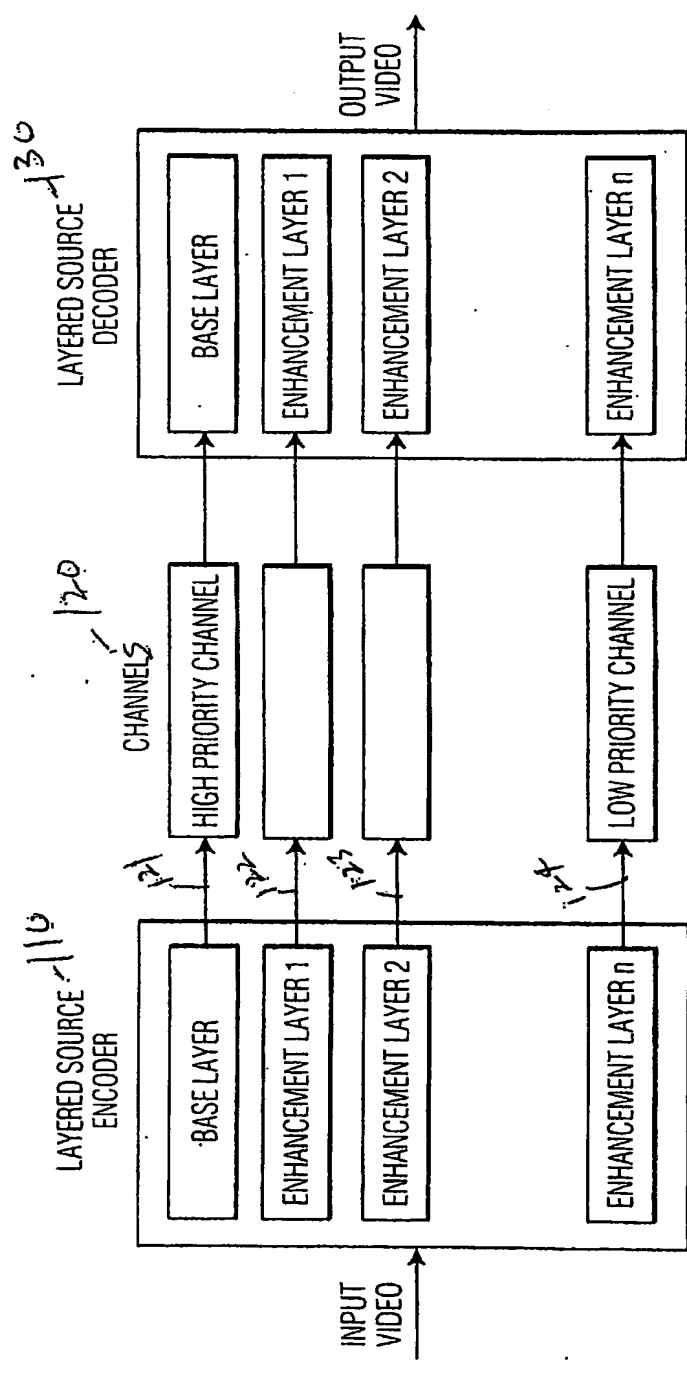


FIG. 3

FIG. 3 is a block diagram of a layered video source encoder and decoder system.

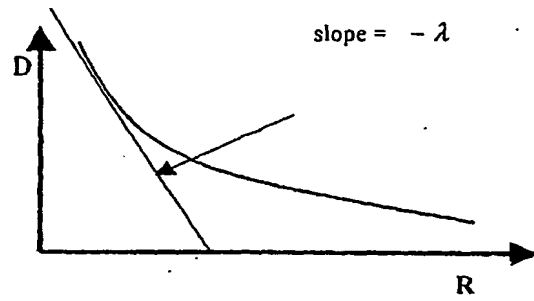


Figure 2: R-D curve.

Fig. 4

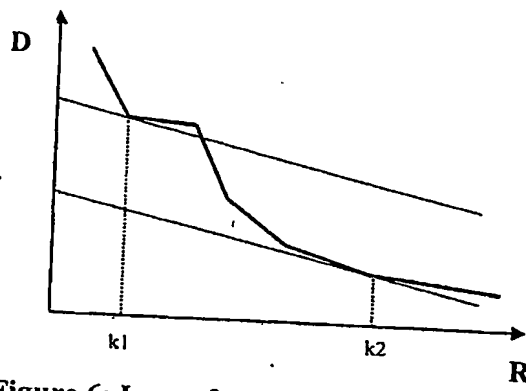


Figure 6: Loss of near optimality in RDDP

FIG. 5

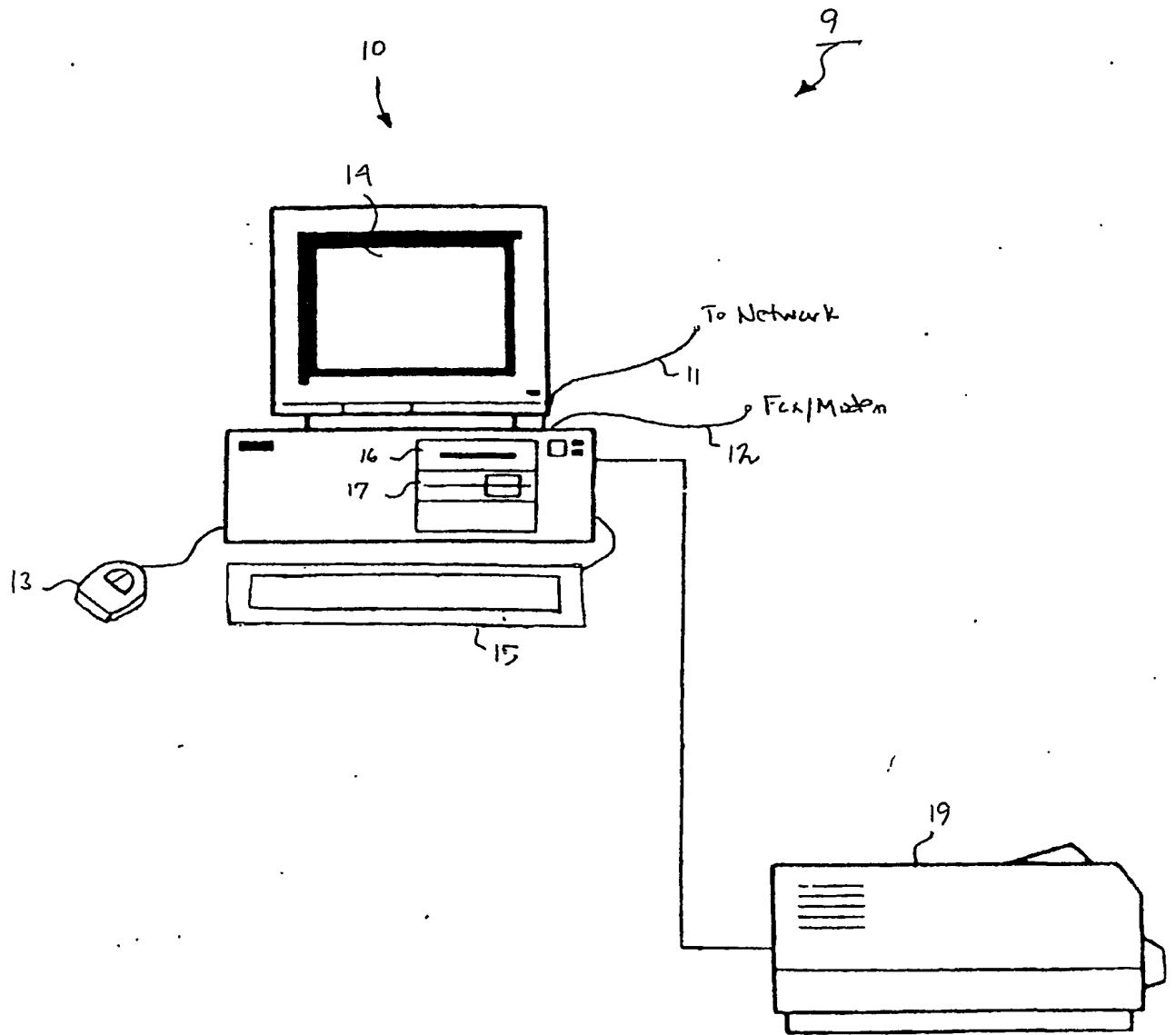
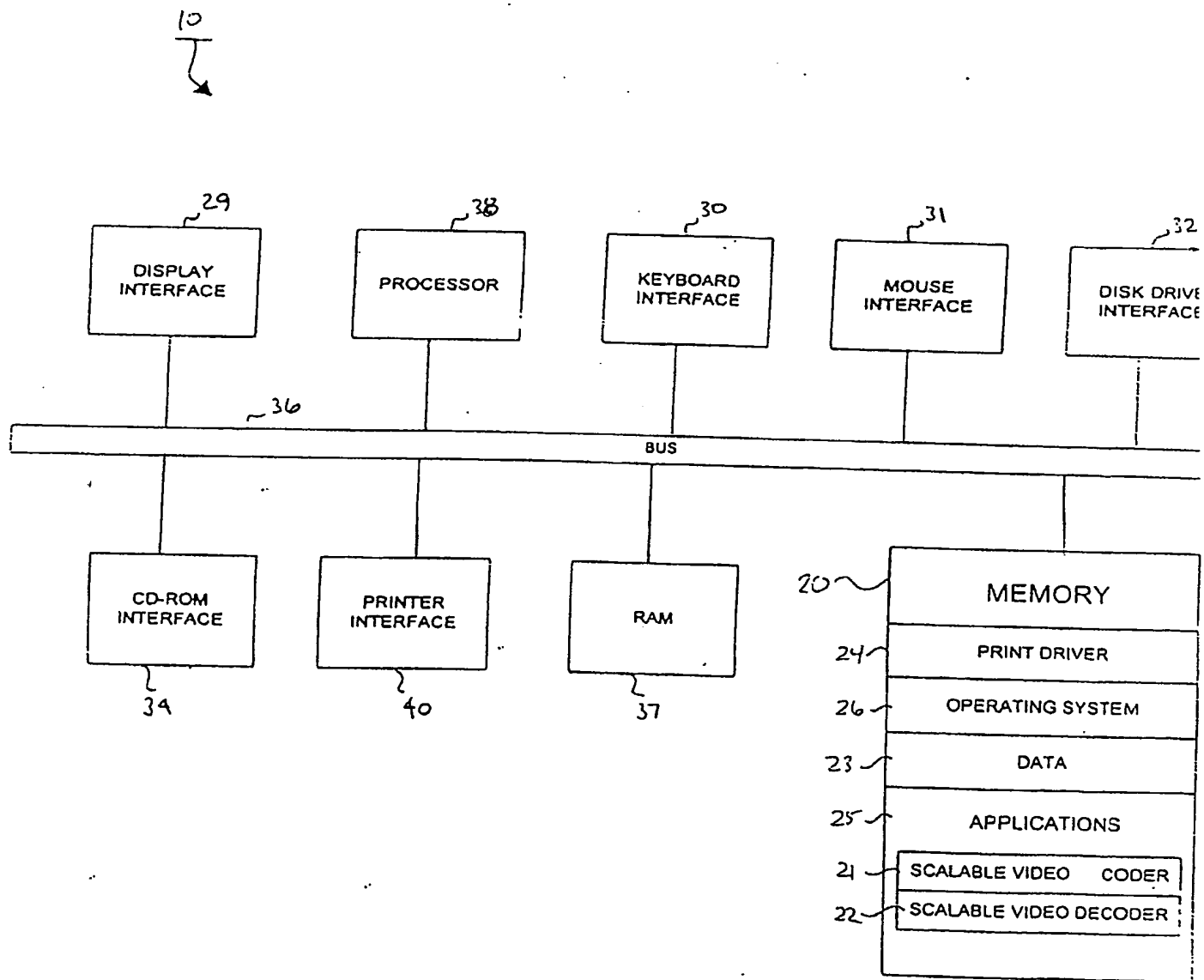


FIG. 6

**FIG. 1**

**This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record.**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**